A Meta-Analysis of the Effectiveness of Treatment for Sexual Offenders: Risk, Need, and Responsivity

2009-01

R. Karl Hanson
Guy Bourgon
Leslie Helmus
Shannon Hodgson

Public Safety Canada

© Her Majesty the Queen in Right of Canada, 2009

Cat. No.: PS3-1/2009-2E-PDF
ISBN No.: 978-1-100-11957-1

Author Note

We would like to thank Caleb Lloyd for coding studies for interrater reliability, Kelly Babchishin for assisting with references, graphs, and formatting, and Annette Hegel for assisting with the coding of German-language studies. We would also like to thank the following individuals for providing information (e.g., unpublished manuscripts, program manuals, additional data) that was helpful in the coding of studies: Jayne Allam, Howard Barbaree, Chuck Borduin, Jackie Craissati, Art Gordon, Heather Gretton, Andrew Haag, Leigh Harkins, Stephen Huot, Stephen Kramer, Dan Kriegman, Ian Lambie, Richard Laws, Roxanne Lieb, Bill Lindsay, Duyen Luong, Bruce Malcolm, Janice Marques, Bill Marshall, Bob McGrath, Timothy McGuire, Michael Miner, Mark Olver, Vern Quinsey, Lorraine Reitzel, Marnie Rice, Frans Ruddijs, Fred Schmidt, Jane Silovsky, Michael Smith, Joe Stommel, Linda Williams, Robin Wilson, Wanda Wishart, and Jim Worling.

The views expressed are those of the authors and not necessarily those of Public Safety Canada.

Correspondence concerning this article should be addressed to R. Karl Hanson, Corrections Research, Public Safety Canada, 340 Laurier Ave., West, Ottawa, Ontario, Canada, K1A 0P8. E-mail: Karl.Hanson@ps-sp.gc.ca

Abstract

The effectiveness of treatment for sexual offenders remains controversial, even though it is widely agreed that certain forms of human service interventions reduce the recidivism rates of general offenders. The current review examined whether the principles associated with effective treatments for general offenders (Risk–Need–Responsivity: RNR) also apply to sexual offender treatment. Based on a meta-analysis of 23 recidivism outcome studies meeting basic criteria for study quality, the unweighted sexual and general recidivism rates for the treated sexual offenders were lower than the rates observed for the comparison groups (10.9% [n = 3,121] versus 19.2% [n = 3,625] for sexual recidivism; 31.8% [n = 1,979] versus 48.3% [n = 2,822] for any recidivism). Programs that adhered to the RNR principles showed the largest reductions in sexual and general recidivism. Given the consistency of the current findings with the general offender rehabilitation literature, we believe that the RNR principles should be a major consideration in the design and implementation of treatment programs for sexual offenders.

Table of Contents

Abstract i

Introduction 1

Method 3

Selection of Studies 3

Measures and Coding Procedure 4

CODC Study Quality Guidelines 4

Adherence to Risk, Need, and Responsivity 9

Study Descriptors 9

Summary of 23 Accepted Studies 10

Index of Treatment Effectiveness 11

Summarizing Findings 11

Results 13

Discussion 23

Implications for Treatment Providers 25

Implications for Researchers 25

References 27

List of Tables and Figure

Table 1	Studies Included In Meta-Analysis	5
Table 2	Studies: Results	7
Table 3	Meta-Analysis of the Effects of Treatment on Recidivism	14
Table 4	Effectiveness of Treatment According to Adherence to Principles of Risk, Need, and Responsivity	18
Figure 1	Treatment Effectiveness by Year and Adherence to RNR Principles	21

Introduction

Does treatment work for sexual offenders? The debate in the scientific literature remains divided: some reviews have concluded that psychological treatment reduces the recidivism risk of sexual offenders (Gallagher, Wilson, Hirschfield, Coggeshall, & MacKenzie, 1999; Hall, 1995; Hanson et al., 2002; Lösel & Schmucker, 2005) whereas other reviews have concluded that the evidence is insufficient to make such a conclusion (Furby, Weinrott, & Blackshaw, 1989; Harris, Rice, & Quinsey, 1998; Kenworthy, Adams, Brooks-Gordon, & Fenton, 2004; Rice & Harris, 2003).

The largest review is the meta-analysis conducted by Lösel and Schmucker (2005), who combined 69 studies to compare the recidivism rates of 9,512 treated sexual offenders to 12,669 untreated sexual offenders. They concluded that there was a positive treatment effect on sexual and other recidivism, and that cognitive-behavioural programs were more effective than other psychosocial approaches. In contrast, Kenworthy et al.'s (2004) review of nine random-assignment studies concluded that "the ethics of providing this still-experimental treatment to a vulnerable and potentially dangerous group of people outside of a well-designed evaluative study are debatable" (p. 2).

All reviews have concluded that more and better studies are needed. Few studies have used strong research designs (i.e., random assignment) and there are even fewer studies with strong research designs examining interventions consistent with contemporary standards. Consequently, reviewers are forced to consider which of the less-than-ideal studies are "good enough." Often they disagree. The "best" studies identified in the reviews of Rice and Harris (2003), Kenworthy et al. (2004), Hanson et al. (2002), and Lösel and Schmucker (2005) were different. Only one study was included in all four lists: California's Sex Offender Treatment and Evaluation Project (SOTEP; Marques, Wiederanders, Day, Nelson, & van Ommeren, 2005). The SOTEP study is unique in that it used a strong research design (random assignment) to evaluate a credible (i.e., cognitive-behavioural) treatment program for adult sexual offenders. The SOTEP study found that the relapse prevention treatment examined was not effective in reducing recidivism; in contrast, a large number of studies using weaker designs have found that similar treatments were associated with reductions in both sexual and general recidivism (Hanson et al., 2002; Lösel & Schmucker, 2005).

Rating study quality is a complex judgement, and, like most complex judgements, tends to be most reliable when the ratings are based on explicit rules. A large number of scales and checklists have been developed within the medical field to assess the quality of randomized and clinical trials (Juni, Witschi, Bloch, & Egger, 1999; Moher et al., 1995). In criminology, where random assignment studies are comparatively rare, the Maryland scale (Sherman et al., 1997) is one of the most commonly used (see also Aos, Phipps, Barnoski, & Liebe, 1999; Lösel & Schmucker, 2005). The Maryland scale is not an ideal measure for the purpose of meta-analysis, however, as it combines concerns about statistical power with concerns about bias. The Maryland scale assumes that the reviewers are interested in the conclusions of the different studies rather than aggregating the data through secondary analysis.

The current review used the Guidelines of the Collaborative Outcome Data Committee (CODC, 2007a, 2007b) to rate study quality. These Guidelines were explicitly developed to rate the

quality of sexual offender treatment outcome studies in the context of meta-analysis (for a summary, see Helmus, 2008). Subsequent research has extended the CODC Guidelines to other domains, such as evaluations of drug treatment courts (Gutierrez, 2008) and community supervision (Simpson et al., 2008). The CODC Guidelines express the consensus opinion of experts concerning the features that increase or decrease the confidence that the results are an unbiased estimate of treatment effectiveness. As such, they provide a credible and explicit method of rating study quality that can be applied to a wide range of studies in the criminal justice field, including both random and non-random assignment studies.

Non-random assignment studies can yield results that are the same as high quality, random assignment studies; alternately, non-random assignment studies can be systematically biased, such that they overestimate or underestimate treatment effects (Kunz & Oxman, 1998). Until a sufficient number of high quality, random assignment studies are completed, it is impossible to predict the degree of similarity between the findings of random assignment studies and non-random assignment studies of sexual offender treatment outcome. However, evidence of meaningful and predictable variations in the findings based on the quality of the interventions would support the position that the results of the available studies are determined by more than design artefacts.

Determining treatment quality for sexual offenders is difficult given the ambiguity in the research data. One approach to making theoretically grounded decisions concerning the quality of sexual offender treatment programs is to consider sexual offender treatment a special case of the treatment of general offenders. Although at one time there was widespread scepticism concerning the effectiveness of any correctional program (Martinson, 1974), much is now known about the interventions that are likely to reduce the recidivism rates of general offenders. The evidence supporting effective treatment for general offenders is strong, including large numbers of studies – many of the highest methodological rigour (see Andrews & Bonta, 2006, chap. 10; D. B. Wilson, Bouffard, & Mackenzie, 2005).

The human service interventions that are most effective for general offenders are those that follow the principles of Risk, Need, and Responsivity (Andrews, Bonta, & Hoge, 1990; Bonta & Andrews, 2007). Briefly stated, treatments are most likely to be effective when they treat offenders who are likely to reoffend (moderate or higher risk), target characteristics that are related to reoffending (criminogenic needs), and match treatment to the offenders' learning styles and abilities (responsivity, cognitive-behavioural interventions work best). These results have been observed for high quality random assignment studies and non-random assignment studies, and the basic pattern of results has been replicated through meta-analyses by independent groups (Andrews & Bonta, 2006; Landenberger & Lipsey, 2005; D. B. Wilson et al., 2005).

The primary question addressed in the current meta-analysis is whether the principles of effective interventions for general offenders also apply to psychological treatment of sexual offenders. A secondary aim was to assess sex offender treatment effectiveness using only studies that met a minimum level of study quality (assessed using the CODC Guidelines). The studies examined compared the recidivism rates of a group of treated sexual offenders to an untreated comparison group. The meta-analysis originally planned to assess both psychological and medical interventions (e.g., hormone treatments) but none of the surgery or drug studies met the minimum level of study quality established by the CODC Guidelines.

Method

Selection of Studies

Computer searches of PsychINFO, Web of Science, Digital Dissertations, and the National Criminal Justice Reference System (NCJRS) were conducted using the following key terms: sex* offend*, rape, rapist*, child molest*, pedophil*, paedophil*, exhibitionis*, sexual assault, incest, voyeur*, frotteur*, indecent exposure, sexual* devian*, paraphilia*, and treatment, outcome, intervention, program, recid*, reoffen*, relapse, or failure. Additional articles were sought through an examination of the reference lists of the collected articles and of review articles in this area.

Identifying studies for inclusion in the meta-analysis occurred in two phases. The first phase involved identifying studies that met minimum criteria to be considered a treatment outcome study. A study had to examine the effectiveness of treatment by comparing the recidivism rates (sexual, violent, or general) of a sample of treated sex offenders with a comparison group of sex offenders. Sex offenders were defined as offenders with sexually motivated offences against an identifiable victim ("Category A" offences – see Harris, Phenix, Hanson, & Thornton, 2003). Studies involving illegal but consensual sexual crimes (e.g., prostitution/John school) were excluded. The comparison group could have received no treatment, an alternate treatment, or less treatment. As well, we included studies that compared the recidivism rate of a treated sample to established norms. Studies were excluded in which the only difference between groups was a form of supervision. Additionally, studies that included recidivism rates for different groups (including a treatment group) but did not intend to test the effectiveness of treatment were also excluded. Two studies (Abracen, Looman, & Nicholaichuk, 1999; Mander, Atrops, Barnes, & Munafo, 1996) were not included following personal communications from the authors indicating unresolved anomalies in their data (A. R. Barnes, November 17, 1999; J. Looman, February 7, 2000).

Different articles that reported findings based on the same sample of offenders (or overlapping samples) were considered as one study. When possible, the report with the largest sample size and the longest follow-up was chosen as the primary study. If overlapping articles had different research designs, the strongest research design (assessed using the CODC Guidelines) was chosen. Information was often drawn from different sources (e.g., recidivism information from a journal article, and the program description from an unpublished internal report). A list of all sources used to code each study is available upon request. As of May, 2008, our search yielded a total of 130 distinct and usable studies of treatment for sex offenders.

In the second phase of study inclusion, studies had to meet a minimum level of study quality. Each of the 130 studies identified in the first phase were rated on the Collaborative Outcome Data Committee (CODC) Study Quality Guidelines (described in further detail below). Final ratings on the CODC Guidelines include four categories: rejected, weak, good, and strong. Of the 130 studies, 105 were rated as rejected, 19 as weak, 5 as good, and 1 as strong. The rejected studies were not considered further but a list of these studies is available upon request.

Two additional studies were excluded because they addressed unique research questions, leaving a total of 23 studies available for analysis. Of these two studies, one (Craissati & Beech, 2005) compared the same program delivered in two different formats (individual versus group). For this study (rated weak), we could not form a hypothesis on the direction of a treatment effect. The second study removed was the

only study with a strong CODC rating as well as the only study that examined children with sexually intrusive behaviour problems (Carpentier, Silovsky, & Chaffin, 2006). Given that interventions that are effective for children are unlikely to apply to older persons, the review restricted itself to studies of adolescent (k = 4) or adult sexual offenders (k = 19; see Table 1). The review was also restricted to psychological interventions as none of the studies of medical interventions met the minimum criteria for study quality. Additionally, none of the studies that compared the recidivism rates of a treatment group to established norms met the minimum criteria for study quality.

Measures and Coding Procedure

CODC Study Quality Guidelines. A committee of 12 experts in the area of sex offender research developed the Collaborative Outcome Data Committee Guidelines for the evaluation of research on sex offender treatment outcome (CODC, 2007a, 2007b). The CODC Guidelines contain 20 items (a 21st item is included for cross-institutional designs) organized into the following seven categories: administrative control of independent variables, experimenter expectancies, sample size, attrition (drop-out rates), equivalency of groups, outcome variables, and correct comparisons conducted. The items concern the extent to which the study's features introduce bias in the estimation of the treatment effect, or influence the confidence that can be placed in the study's findings.

After rating the individual items, the evaluator forms a global structured judgement as to the extent of "bias" inherent in the research design (coded as either minimal bias, some bias, or considerable bias). The direction of bias is also coded (as increasing or decreasing the treatment effect, or unknown direction), and the "confidence" that can be placed in the study's findings (coded as little confidence, some confidence, or confidence in the results as reported). An explicit method is used to combine the overall judgements of bias and confidence into one of the following four overall study quality categories: strong, good, weak, and rejected.

Two undergraduate students (3rd year Criminology and 4th year Psychology) coded the CODC Guidelines.[1] The authors of the original study were contacted for clarification when insufficient information was available concerning crucial variables (e.g., method of subject assignment to treatment). After rating each study separately, the two raters generated a consensus rating, which was reviewed by either or both of the first two authors. Whenever possible, data were coded to fit the research design that minimized the likelihood of pre-existing differences between the treatment and comparison groups. Consequently, the recidivism rates reported in Table 2 do not always correspond to those reported in the original articles. For example, if sufficient information was available, drop-outs were included in the treatment group. If survival data were presented, recidivism rates were taken directly from the graph for a standard follow-up period.

The student raters received five days of training on the use of the CODC Guidelines. This training primarily involved rating and reviewing eight practice studies with a trainer.[2] The two raters then independently coded 10 studies, representing a variety of study designs and publication types. The raters completely agreed on the basic method of subject assignment (e.g., random assignment, retrospective cohort, risk band/norm designs). On the Global rating of study quality, the coders agreed on nine of the 10 studies (*ICC* = .95). There was 100% agreement on Global Confidence (Kappa = 1.0; *ICC* = 1.0),

[1] Leslie Helmus and Shannon Hodgson, respectively
[2] Guy Bourgon

Table 1

Studies Included in Meta-Analysis

Study	Design	CODC rating	Age group	Adherence to RNR principles			
				Risk	Need	Responsivity	Total
Bakker et al. (1998)	Retrospective Cohort	Weak	Adults	No	Yes	Yes	2
Borduin et al. (1990)	Randomized Trial	Weak	Adolescents	Yes	Yes	Yes	3
Borduin et al. (2009)	Randomized Trial	Good	Adolescents	Yes	Yes	Yes	3
Cooper (2000)	Other Concurrent Comparison Group	Weak	Adolescents	No	No	Yes	1
Davidson (1984)	Retrospective Cohort	Weak	Adults	Yes	No	Yes	2
Friendship et al. (2003)	Need, Volunteer, & Dropout[a]	Weak	Adults	No	Yes	Yes	2
Hanson et al. (2004)	Other Concurrent Comparison Group	Good	Adults	No	No	No	0
Hanson et al. (1993)	Retrospective Cohort	Weak	Adults	No	No	Yes	1
Harkins (2004)	Need, Volunteer, & Dropout[a]	Weak	Adults	Yes	Yes	Yes	3
Marques et al. (2005)	Randomized Trial	Good	Adults	No	Yes	Yes	2
L. E. Marshall et al. (2008)	Need, Volunteer, & Dropout[a]	Weak	Adults	No	No	Yes	1
W. L. Marshall et al. (1991)	Retrospective Cohort	Weak	Adults	Yes	No	Yes	2
Martin (1998)	Retrospective Cohort	Weak	Adults	No	Yes	Yes	2
McGrath et al. (1998)	Need, Volunteer, & Dropout[a]	Good	Adults	No	Yes	Yes	2
Meyer & Romero (1980)	Researcher Assigned Non-Randomized	Good	Adults	No	No	No	0

Table 1 continues

Table 1 (*Continued*)

Study	Design	CODC rating	Age Group	Adherence to RNR principles			
				Risk	Need	Responsivity	Total
Nathan et al. (2003)	Other Concurrent Comparison Group	Weak	Adults	No	Yes	Yes	2
Procter (1996)	Retrospective Cohort	Weak	Adults	No	No	Yes	1
Robinson (1995)	Randomized Trial	Weak	Adults	Yes	Yes	Yes	3
Ruddijs & Timmerman (2000)	Other Concurrent Comparison Group	Weak	Adults	No	No	No	0
Taylor (2000)	Other Concurrent Comparison Group	Weak	Adults	Yes	No	No	1
Ternowski (2004)	Need, Volunteer, & Dropout[a]	Weak	Adults	No	No	Yes	1
Wilson et al. (2005)	Need, Volunteer, & Dropout[a]	Weak	Adults	Yes	No	Yes	2
Worling & Curwen (1998)	Other Concurrent Comparison Group	Weak	Adolescents	No	No	Yes	1

[a] In "Need, volunteer & drop-out" designs, the factors that excluded offenders from the treatment group were not fully known, but could reasonably be expected to have involved decisions made by the offenders and treatment providers. Studies that explicitly compared treatment completers to drop-outs were Rejected (see CODC Guidelines, 2007b).

Table 2

Studies: Results

Study	Sample Size		Follow-Up (Years)	Sexual Recidivism		Odds Ratio (Sexual)	Violent Recidivism		Odds Ratio (Violent)	Any Recidivism		Odds Ratio (Any)
	Rx	CG		Rx	CG		Rx	CG		Rx	CG	
Bakker et al. (1998)	238	283	6	.10	.22	.405						
Borduin et al. (1990)	8	8	3	.12	.75	.077	.12	.88	.040	.25	.88	.077
Borduin et al. (2009)	24	24	9	.08	.46	.130				.29	.75	.151
Cooper (2000)	64	25	7	.08	.04	1.510	.23	.24	.939			
Davidson (1984)	101	101	3	.22	.10	2.466	.24	.19	1.337	.37	.34	1.138
Friendship et al. (2003)	647	1910	2	.03	.03	.943	.05	.08	.558	.13	.16	.779
Hanson et al. (2004)	403	321	7	.14[a]	.14	1.000	.31[a]	.31	.990	.47[a]	.50	.900
Hanson et al. (1993)	106	31	21	.33	.39	.786	.42	.42	1.042	.57	.58	.959
Harkins (2004)	53	53	5	.06	.11	.506				.36	.45	.681
Marques et al. (2005)	259	225	8	.19[a]	.20	.965	.38	.32	1.340			
L. E. Marshall et al. (2008)	94	83[b]	3	.01	.05	.283	.01	.08	.164	.04	.13	.313
W. L. Marshall et al. (1991)	17	23		.24	.35	.608						
Martin (1998)	65	56	4	.06	.20	.290				.12	.41	.211

Table 2 continues

Table 2 (*Continued*)

Study	Sample Size		Follow-Up (Years)	Sexual Recidivism		Odds Ratio (Sexual)	Violent Recidivism		Odds Ratio (Violent)	Any Recidivism		Odds Ratio (Any)
	Rx	CG		Rx	CG		Rx	CG		Rx	CG	
McGrath et al. (1998)	71	32	5	.01	.16	.106						
Meyer & Romero (1980)	148	83	-	.14	.07	1.902				.55	.60	.823
Nathan et al. (2003)	65	70	5	.05	.08	.556						
Procter (1996)	54	54	5	.07	.17	.427						
Robinson (1995)	189	46	1							.30	.46	.515
Ruddijs & Timmerman (2000)	56	56	-	.05	.02	2.421						
Taylor (2000)	283	38	7	.17	.24	.624	.37	.34	1.100			
Ternowski (2004)	224	43	-	.07[a]	.14	.436						
Wilson et al. (2005)	60	60	4	.05	.17	.293	.15	.35	.339	.28	.43	.524
Worling & Curwen (1998)	85	46	6	.12	.13	.867				.39	.59	.452

Note: "Rx" indicates treatment group and "CG" indicates comparison group.

[a] Adjusted recidivism rates for the treatment group. When an odds ratio was coded that controlled for other variables, the odds ratio and the unadjusted recidivism rate for the comparison group was used to calculate the adjusted recidivism rate for the treatment group.
[b] A preliminary, unpublished report indicated that 8 members of the comparison group were never released and 3 had been deported, reducing the at-risk sample from 94 to 83 (Malcolm, Marshall, & Marshall, 2004).

90% agreement on Global Bias (*ICC* = .69; Kappa could not be computed), and 70% agreement on Global Direction of Bias (Kappa could not be computed). The level of agreement for the individual items was also high. For most of the categories, the median level of agreement was 1.0.

Adherence to Risk, Need, and Responsivity. The first two authors coded the extent to which each of the 23 accepted studies adhered to the Risk-Need-Responsivity (RNR) principles of effective correctional interventions (Bonta & Andrews, 2007). Programs adhered to the Risk principle when they provided intensive interventions to higher risk offenders and little or no service to low risk offenders. In practice, however, no single study had differentiated treatment services. Studies were therefore coded as adhering to the Risk principle if their treatment group was higher risk than average offenders. Adherence to the Need principle was met if the majority of the treatment targets were significantly related to sexual or general recidivism in previous meta-analytic reviews (Andrews & Bonta, 2006, chap. 9; Gendreau, Little, & Goggin, 1996; Hanson & Morton-Bourgon, 2004, 2005). For sexual offence recidivism, the main criminogenic needs were sexual deviancy, antisocial orientation, sexual attitudes, and intimacy deficits. Examples of non-criminogenic needs are denial, low victim empathy, and social skills deficits (Hanson & Morton-Bourgon, 2004, 2005). Treatment services were considered to meet the Responsivity principle when they provided treatment in a manner and style matched to the learning style of the clients. For offenders, such programs are typically cognitive-behavioural programs run by pro-social therapists skilled at developing respectful ("firm but fair") relationships.

RNR ratings were based on all available information, including program manuals, research articles, reports of accreditation panels, and, in some cases, site visits. The program's adherence to each principle (Risk, Need, and Responsivity) was rated as a yes/no dichotomy. The overall score for treatment adherence ranged from 0-3, representing the number of principles of effective correctional treatment that were followed.

Rater reliability was assessed with a graduate student in criminal psychology with no specific background in sexual offender research.[3] He was trained using seven studies and then independently coded adherence to RNR in 16 studies. The RNR ratings were based on program descriptions drawn from published articles, program manuals, internal reviews, and accreditation reports. All information about the effectiveness of the program was removed prior to reliability coding, with the exception of information concerning the drop-out rate (a useful indicator of Responsivity).

The reliability was good for the rating of adherence to the principles of Risk (Kappa = .73, 88% agreement), and Responsivity (Kappa = .82, 94% agreement), but only fair for rating of the Need principle (Kappa = .42; 75% agreement). Nevertheless, the reliability of the overall rating of adherence to the RNR principles was good (*ICC* = .80).

Study descriptors. For the 23 accepted studies, effect sizes and study descriptors were coded using a standardized coding manual (containing approximately 100 descriptive variables, such as sample size, follow-up period, methods of treatment used). Each study was coded separately by two raters, who then generated a consensus rating to be approved by either or both of the first two authors. The primary coders were the same two undergraduate students who coded study quality (CODC Guidelines). Coding of effect sizes and study descriptors was conducted blind to the ratings of treatment quality.

---

[3] Caleb Lloyd

To examine the reliability of the ratings of effect size and descriptive variables, 16 files were coded by a third rater (the same individual who rated RNR adherence). These ratings were completed after coding the RNR reliability ratings on the blind files. In these 16 files, the original raters identified 42 effect sizes and the reliability rater identified 43, of which there was 87.1% agreement (74/85) (three effects were missing due to errors in preparing the reliability files). There was high agreement on the commonly identified effect sizes (*ICC* = .97; $n = 37$).

Only variables coded as non-missing in at least 10 studies were analyzed in the reliability study. Reliability for the descriptive variables were calculated as *ICC* ($n = 13$), Kappa ($n = 20$) or percent agreement when Kappa could not be calculated ($n = 11$). The Intraclass Correlation Coefficients ranged from .95 to 1.0, with a median value of .99. Kappa values ranged from .22 to 1.0, with a median value of .87. Four variables had Kappa values less than .50. For two questions (Any females? Only adults?), the low Kappa were the result of one rater identifying low base-rate information overlooked by the other rater; these variables were retained. The other two questions were eliminated because of low reliability (Was a treatment manual used? Number of therapists?). Percent agreement ranged from 62% to 94%, with a median value of 87%. The lowest agreement (62%) was for the primary therapeutic orientation (cognitive-behavioural, behavioural, systemic, psychotherapeutic, other, unknown), which increased to 75% agreement (Kappa = .47) for the distinction between cognitive-behavioural interventions and "others." This variable was retained for descriptive purposes only.

Summary of 23 Accepted Studies

Studies were classified as published ($k = 14$) or unpublished ($k = 9$) based on whether the recidivism data used for the meta-analysis were published in a peer-reviewed journal. All other documents were considered "unpublished" (e.g., book chapters, government reports, dissertations).

Most of the studies were based on Canadian ($k = 12$) or American samples ($k = 5$), with three studies from the United Kingdom, two from New Zealand, and one from Holland. The studies were mostly recent (median publication year = 2000, ranging from 1980 to 2009). Total sample size ranged from 16 to 2,557 (median of 135). Most of the studies focused on adult male sex offenders. Four studies specifically examined adolescent sex offenders (see Table 1). Three studies (Borduin, Schaeffer, & Heiblum, 2009; Cooper, 2000; Worling & Curwen, 2000) indicated that their sample contained some female offenders (< 10% of their total sample).

Of the 23 treatment programs, 10 were offered in institutions, 11 in the community, and two in both settings. The major sponsor of the programs was corrections ($k = 16$). One program was a private clinic and the remaining six were "other/unknown." Treatment was delivered between 1965 and 2004, with approximately 90% of the offenders receiving treatment after 1980. Most studies examined specialized treatment programs for sex offenders, although four studies examined the response of sex offenders to programs designed for general offenders (Borduin et al., 2009; Borduin, Henggeler, Blaske, & Stein, 1990; Robinson, 1995; Taylor, 2000).

Recidivism was defined as reconviction in 10 studies and rearrest in 12 studies. In one study, the criterion for recidivism (i.e., reconviction or rearrest) was not specified. The most common source of recidivism information was national criminal justice records ($k = 19$) followed by state/provincial records ($k = 7$). Additional sources of information (e.g., child protection records, self-reports) were used in four studies. These numbers add up to more than 23 because some studies used multiple sources. Twenty-two studies

reported sexual recidivism, 10 studies reported violent (including sexual) recidivism, and 13 studies reported general (any) recidivism. The average follow-up periods ranged from one to 21 years, with a median of 4.7 years.

Index of Treatment Effectiveness

The basic outcome data were 2 x 2 tables containing the recidivism outcomes of the treatment and comparison groups. Following Fleiss (1994), the index selected was the odds ratio, which is defined as the ratio of two ratios: a) the probability of recidivism among the treatment group divided by the probability of non-recidivism in the treatment group (odds recidivism in the treatment group), divided by b) the probability of recidivism in the comparison group divided by the probability of non-recidivism in the comparison group (odds recidivism in the comparison group). Following convention, analyses were conducted on the natural log of the odds ratio because the variance is easily defined (see Fleiss, 1994):

$$LnOR = \ln\left[\frac{(a+.5)(d+.5)}{(b+.5)(c+.5)}\right], \text{ and}$$

$$V_{LnOR} = \frac{1}{a+.5} + \frac{1}{b+.5} + \frac{1}{c+.5} + \frac{1}{d+.5}.$$

LnOR is the natural log of the odds ratio, V_{LnOR} is its variance, and *a*, *b*, *c*, and *d* are the cells of a 2 by 2 table. Note that 0.5 was added to each cell to permit the analysis of tables containing empty cells.

Values of the odds ratio can range from very small (e.g., < .01) to very large (e.g., > 100) with values of 1.0 indicating no difference between the groups. Small values of the odds ratio were coded to indicate treatment effectiveness, i.e., lower recidivism rates in the treatment versus comparison group. When the recidivism base rate is low, the odds ratio approximates the rate ratio. For example, given a base rate of 10%, an odds ratio of .70 can be interpreted as follows: for every 100 untreated sex offenders who recidivate, only 70 treated sex offenders will recidivate.

Odds ratios are a desirable indicator of effect size for describing the relationship between two dichotomous variables because, unlike phi coefficients, they would be expected to remain constant given variation in arbitrary design features, such as the length of follow-up or the relative proportion in the treatment and comparison groups. In contrast, meta-analyses using *r* typically use some form of correction to adjust for the restriction of range expected with low recidivism base rates (common in sexual offender research; Campbell, French, & Gendreau, 2007).

Summarizing Findings

Summary statistics were calculated using both fixed effect and random effect models (Hedges & Vevea, 1998). Each approach asks slightly different questions and neither approach has won universal acceptance (Whitehead, 2002, section 6.3). On a conceptual level, the conclusions of the fixed effect analyses are restricted to the particular set of studies included in the meta-analysis. In contrast, the random effect model aims for conclusions that apply to the population of studies of which the current sample of studies is a part.

In practical terms, the random effect model includes an additional between-study error term representing the unexplained variation across studies (a constant). Compared to the fixed effect model, the random effect model has higher variance estimates (wider confidence intervals), and the differences in sample size across the studies is given less importance. Consequently, the random effects model gives relatively more weight to small studies than does the fixed effect model (approximating unweighted averages).

When the assumptions are violated, the fixed effect model is too liberal and the random effect model is too conservative (Overton, 1998). The results of the random effect and fixed effect models converge as the amount of between-study variability decreases. When the variation between studies is less than would be expected by chance ($Q <$ degrees of freedom, using Cochran's Q statistic; Hedges & Olkin, 1985), both approaches yield identical results.

Fixed effects estimates of means, standard errors, and moderator effects (meta-regression) were calculated using the formula and procedures presented in Hedges (1994). Random effects estimates of means and standard errors were calculated using Formulae 10, 12, and 14 from Hedges and Vevea (1998). Hand calculations or SPSS syntax was used for all analyses, except for the random effect meta-regression, which were computed using Comprehensive Meta-Analysis Version 2.0 (Biostat; Borenstein, Hedges, Higgins, & Rothstein, 2006).

Results

In total, 22 studies examined the rates of sexual recidivism, which comprised 3,121 treated offenders and 3,625 offenders in the comparison groups. The sexual recidivism rate of the treatment groups ranged from 1.1% to 33.3%, with an unweighted mean of 10.9%. The sexual recidivism rate for the comparison groups ranged from 1.8% to 75.0%, with an unweighted mean of 19.2%. In 17 of the 22 studies, the sexual recidivism rate of the treatment group was lower than the recidivism rate of the comparison groups (exact $p = .0085$ against the null hypothesis of $p = .50$, one-tailed). The odds ratio for sexual recidivism ranged from .08 to 2.47 with a fixed-effect weighted mean of .77 (95% confidence interval [C.I.] of .65 to .91). There was, however, more variability than would be expected by chance ($Q = 47.17$, $df = 21$, $p < .001$). The random-effects weighted mean was .66 (95% C.I. of .49 to .89). In both fixed-effect and random-effects analyses, the 95% confidence intervals for the odds ratios do not include 1.0, indicating significantly lower sexual recidivism rates in the treatment groups than in the comparison groups.

There were 10 studies that examined violent (including sexual) recidivism, comprising 2,021 treated offenders and 2,802 offenders in the comparison group. Studies that reported only sexual recidivism or only non-sexual violence were not considered in these analyses. The combined sexual or violent recidivism rate ranged from 1.1% to 43.1% for treatment groups (unweighted mean of 22.9%), and from 8.1% to 87.5% for the comparison groups (unweighted mean of 32.0%). In 6 of the 10 studies, the recidivism rate for the treatment group was lower than the recidivism rate of the comparison group (exact $p = .377$, one-tailed). The odds ratios for violent recidivism ranged from .04 to 1.34 with a fixed-effect weighted mean of .92 (95% C.I. of .78 to 1.07) and significant variability ($Q = 26.63$, $df = 9$, $p < .005$). The random-effects weighted mean was .81 (95% C.I. of .58 to 1.14). In other words, the combined sexual and violent recidivism rates were not significantly lower for the treatment groups relative to the comparison groups.

General (any) recidivism was examined in 13 studies (1,979 treated; 2,822 comparison subjects). Studies that reported only sexual recidivism, violent recidivism, non-sexual recidivism, or non-violent recidivism were excluded from these analyses. The overall (sexual, violent, or non-violent) recidivism rate for the treatment group was lower (unweighted $M = 31.8\%$, range of 4.3% to 56.9%) than the overall recidivism rate of the comparison groups (unweighted $M = 48.3\%$, range of 13.3% to 87.5%). In 12 of the 13 studies, the recidivism rate favoured the treatment group (exact $p = .0017$, one-tailed). The odds ratios for any recidivism ranged from .07 to 1.14 with a fixed-effect mean of .75 (95% C.I. of .66 to .86) and significant variability ($Q = 29.82$, $df = 12$, $p < .005$). The random-effects weighted mean was .61 (95% C.I. of .47 to .80). In other words, in both fixed-effect and random-effects analyses, the recidivism rate was significantly lower in the treatment groups than in the comparison groups.

Using fixed-effect analyses, the treatment effects on sexual recidivism were smaller in the good quality studies than the weak studies (see Table 3). This comparison was not significant, however, using random-effects analysis. No significant differences were noted in the treatment effects for the published or unpublished studies. For the reduction of sexual recidivism, treatment appeared to be equally effective for adults and adolescents, and did not depend on whether the program was delivered in the community or institution (both fixed-effect and random-effect comparisons were not significant).

Table 3

Meta-Analysis of the Effects of Treatment on Recidivism

	Average Odds Ratios (95% Confidence Intervals)		Q/Q_{Δ}			
	Random	Fixed	Random	Fixed	*n*	(*k*)
Effects for Sexual Recidivism						
All Studies	.66 (.49 to .89)	.77 (.65 to .91)		47.17***	6,746	22
Research Design			0.59	5.03*		
Good	.75 (.42 to 1.37)	.94 (.74 to 1.20)		14.09**	1,590	5
Weak	.63 (.45 to .88)	.64 (.51 to .81)		28.05*	5,156	17
Published?			0.08	3.66		
Yes	.70 (.50 to .998)	.86 (.71 to 1.04)		25.91*	4,984	14
No	.63 (.37 to 1.09)	.60 (.44 to .82)		17.60*	1,762	8
Age of Sample			1.06	1.84		
Adults	.71 (.53 to .95)	.79 (.67 to .94)		37.39**	6,462	18
Adolescents	.38 (.10 to 1.41)	.47 (.22 to .98)		7.94*	284	4
Location			0.06	0.44		
Institution	.69 (.48 to .99)	.74 (.59 to .91)		22.39**	5,024	11
Community	.59 (.34 to 1.04)	.83 (.63 to 1.08)		24.34**	1,722	11

Table continues

Table 3 (*Continued*)

	Average Odds Ratios (95% Confidence Intervals)		Q/Q_{Δ}			
	Random	Fixed	Random	Fixed	*n*	(*k*)
Effects for Sexual or Violent Recidivism						
All Studies	.81 (.58 to 1.14)	.92 (.78 to 1.07)		26.63**	4,823	10
Research Design			4.99*	6.84**		
Good	1.11 (.83 to 1.48)	1.08 (.88 to 1.32)		1.77	1,208	2
Weak	.66 (.41 to 1.08)	.70 (.54 to .90)		18.05*	3,615	8
Published?			1.59	1.31		
Yes	.68 (.43 to 1.07)	.88 (.74 to 1.05)		24.98***	4,211	7
No	1.16 (.75 to 1.81)	1.16 (.75 to 1.81)		0.35	612	3
Age of Sample			0.75	0.88		
Adults	.86 (.62 to 1.20)	.93 (.79 to 1.09)		20.48**	4,718	8
Adolescents	.24 (.01 to 5.12)	.58 (.22 to 1.53)		5.27	105	2
Location			0.96	0.08		
Institution	.91 (.59 to 1.43)	.94 (.75 to 1.18)		15.03*	3,874	6
Community	.54 (.22 to 1.28)	.90 (.72 to 1.12)		11.53**	949	4

Table 3 continues

Table 3 (*Continued*)

	Average Odds Ratios (95% Confidence Intervals)		Q/Q_{Δ}			
	Random	Fixed	Random	Fixed	*n*	(*k*)
Effects for General (Any) Recidivism						
All Studies	.61 (.47 to .80)	.75 (.66 to .86)		29.82**	4,801	13
Research Design			0.52	2.85		
Good	.64 (.34 to 1.21)	.85 (.70 to 1.04)		7.81**	1,003	3
Weak	.58 (.42 to .80)	.67 (.56 to .81)		19.16*	3,798	10
Published?			0.04	1.06		
Yes	.64 (.48 to .86)	.77 (.67 to .90)		18.51*	4,137	9
No	.57 (.29 to 1.09)	.64 (.45 to .90)		10.25*	664	4
Age of Sample			7.88**	8.86**		
Adults	.71 (.56 to .90)	.79 (.69 to .90)		17.17*	4,606	10
Adolescents	.24 (.09 to .65)	.31 (.17 to .56)		3.79	195	3
Location			0.14	0.37		
Institution	.64 (.33 to .92)	.72 (.59 to .87)		13.67*	3,531	7
Community	.53 (.33 to .87)	.78 (.65 to .94)		15.78**	1,270	6

Note: *k* is the number of studies.
*$p < .05$, ** $p < .01$, *** $p < .001$

For the outcome of violent (including sexual) recidivism, the effect of treatment was worse in the good studies than in the weak studies according to both fixed-effect and random-effects models. There were no differences in the effects according to whether the studies were published or unpublished, had examined adult or adolescent offenders, or were delivered in the community or an institution.

For the outcome of any recidivism, there were no differences in the effects according to whether the research design was good or weak, whether the study was published or unpublished, or whether the treatment was delivered in the community or an institution. Treatment appeared to be more effective for adolescent offenders than for adult offenders according to both the random-effects and fixed-effects models. This difference was due to large effects on general recidivism found in two studies of Multi-Systemic Therapy, a treatment specifically designed for general recidivism among adolescent offenders (Borduin et al., 1990; Borduin et al., 2009).

The next set of analyses examined the effectiveness of treatment according to adherence to the principles of Risk, Need, and Responsivity (see Table 4). For the outcome of sexual recidivism (22 studies), fixed-effects analyses found that programs were more effective if they targeted criminogenic needs (Need principle) and delivered in a manner that was likely to engage the offenders (Responsivity principle). Both the fixed-effects and random-effects models found that effectiveness of treatment increased according to the total number of principles adhered to (none, only one, any two, all three, corresponding to odds ratios of 1.17, .64, .63 and .21, respectively, using random-effects estimates). The odds ratios for the high risk samples were not significantly different than odds ratios for the other samples, although the direction of the effect was consistent with the Risk principle (stronger treatment effects for the high risk offenders).

For the 10 studies that examined sexual and violent recidivism as the outcome variable, there were no significant differences based on adherence to the risk, need, responsivity principles (although all effects were in the expected directions).

For the 13 studies that examined general (any) recidivism, the fixed-effect model found stronger treatment effects for treatments adhering to the Responsivity principle, as well as for the total number of principles adhered to. All the effects were in the direction predicted by the RNR principles, although none were statistically significant using the random-effects model.

Recent treatments were more effective, on average, than the treatments delivered in previous decades (see Figure 1). The starting date for the treatment ranged between 1965 and 1997 (M = 1986, SD = 8.5 years, median of 1989). For all outcomes, the linear association was statistically significant for both the fixed-effects and random-effects models. (For fixed effects, sexual recidivism: b = -.042, Z = 3.62, $p < .001$; violent recidivism, b = - .038, Z = 2.93, p = .0034; any recidivism, b = -.020, Z = 2.44, p = .015. For random effects, sexual recidivism: b = -.042, Z = 3.00, p = .003; violent recidivism, b = - .037, Z = 2.54, p = .011; any recidivism, b = -.025, Z = 2.24, p = .025).

Table 4

Effectiveness of Treatment According to Adherence to Principles of Risk, Need, and Responsivity

	Average Odds Ratios (95% Confidence Intervals)		Q/Q_{Δ}			
	Random	Fixed	Random	Fixed	*n*	(*k*)
Effects for Sexual Recidivism						
High Risk Offenders			.41	.35		
Yes	.48 (.21 to 1.11)	.69 (.45 to 1.05)		20.13**	853	7
No	.72 (.53 to .97)	.79 (.66 to .95)		26.69*	5,893	15
Targets Criminogenic Needs			3.70	4.87*		
Yes	.45 (.27 to .75)	.63 (.49 to .81)		22.31**	4,091	9
No	.86 (.60 to 1.21)	.92 (.73 to 1.15)		19.99	2,655	13
Responsivity			3.37	5.26*		
Yes	.57 (.40 to .80)	.67 (.55 to .83)		37.93**	5,358	18
No	1.05 (.69 to 1.60)	1.02 (.76 to 1.36)		3.98	1,388	4
Total RNR Principles			5.50*	6.83**		
None	1.17 (.77 to 1.77)	1.10 (.81 to 1.50)		2.28	1,067	3
One	.64 (.42 to .92)	.64 (.42 to .92)		3.20	1,226	7
Two	.63 (.38 to 1.08)	.74 (.58 to .93)		25.93**	4,283	9
All Three	.21 (.070 to .64)	.22 (.089 to .57)		2.71	170	3

Table 4 continues

Table 4 (*Continued*)

	Average Odds Ratios (95% Confidence Intervals)		Q/Q_{Δ}			
	Random	Fixed	Random	Fixed	*n*	(*k*)
Effects for Sexual or Violent Recidivism						
High Risk Offenders			0.23	0.30		
Yes	.59 (.23 to 1.53)	.82 (.54 to 1.25)		12.39**	659	4
No	.87 (.61 to 1.26)	.93 (.79 to 1.11)		13.95*	4,164	6
Targets Criminogenic Needs			0.14	0.38		
Yes	.61 (.23 to 1.60)	.85 (.65 to 1.12)		15.76***	3,057	3
No	.88 (.61 to 1.25)	.95 (.78 to 1.16)		10.49	1,766	7
Responsivity			0.79	1.16		
Yes	.69 (.41 to 1.15)	.84 (.67 to 1.05)		25.39***	3,778	8
No	1.00 (.80 to 1.26)	1.00 (.80 to 1.26)		0.08	1,045	2
Total RNR Principles			1.00	1.39		
None	.99 (.78 to 1.26)	.99 (.78 to 1.26)		-	724	1
One	.89 (.51 to 1.53)	.93 (.59 to 1.46)		3.96	720	4
Two	.80 (.43 to 1.47)	.87 (.68 to 1.11)		15.95**	3,363	4
All Three	.04 (.00 to .48)	.04 (.00 to .48)		-	16	1

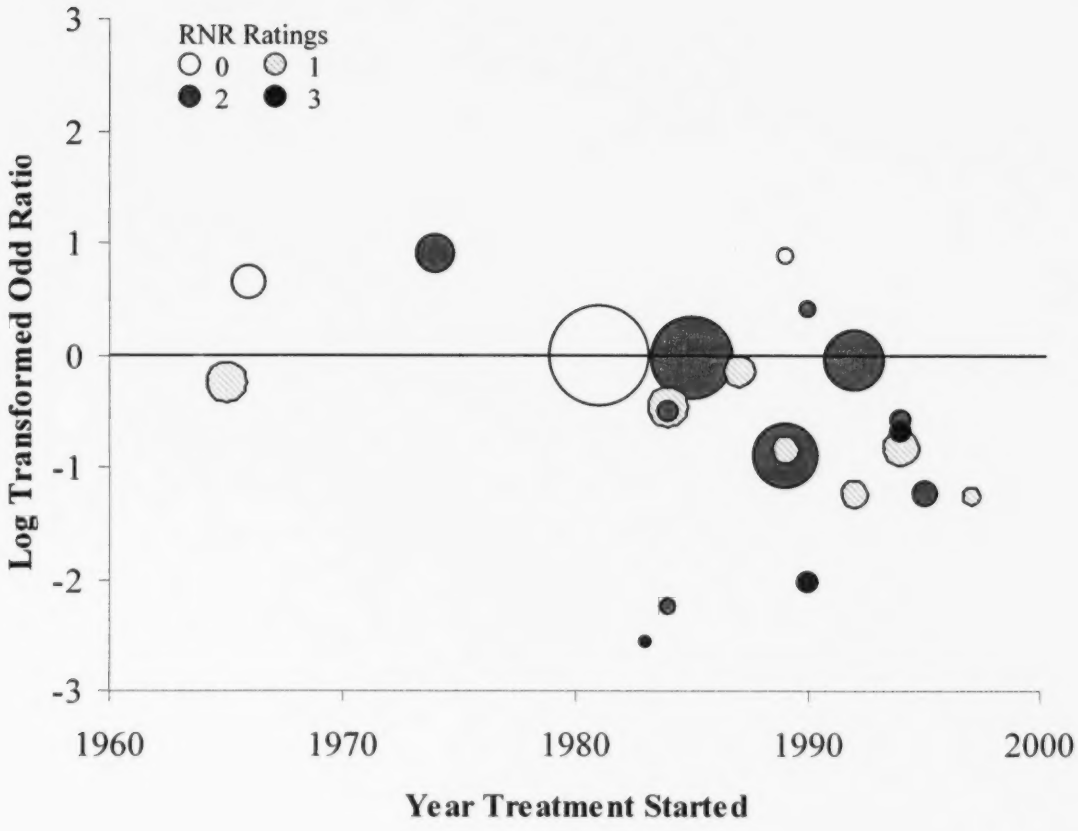
Table 4 continues

Table 4 (*Continued*)

	Average Odds Ratios (95% Confidence Intervals)		Q/Q_{Δ}			
	Random	Fixed	Random	Fixed	*n*	(*k*)
Effects for General (Any) Recidivism						
High Risk Offenders			0.36	1.79		
Yes	.51 (.29 to .90)	.62 (.45 to .85)		13.09*	727	6
No	.67 (.50 to .90)	.78 (.68 to .91)		14.94*	4,074	7
Targets Criminogenic Needs			2.96	3.79		
Yes	.40 (.23 to .72)	.63 (.51 to .79)		17.22**	3,083	6
No	.78 (.60 to 1.01)	.83 (.70 to .99)		8.81	1,718	7
Responsivity			2.86	5.05*		
Yes	.53 (.37 to .75)	.65 (.54 to .78)		24.68*	3,846	11
No	.89 (.73 to 1.09)	.89 (.73 to 1.09)		0.09	955	2
Total RNR Principles			3.35	5.63*		
None	.89 (.73 to 1.09)	.89 (.73 to 1.09)		0.09	955	2
One	.55 (.30 to 1.01)	.56 (.34 to .91)		3.07	441	3
Two	.62 (.36 to 1.07)	.74 (.59 to .91)		10.72*	3,000	4
All Three	.3€ (.17 to .78)	.45 (.28 to .70)		6.58	405	4

Note: k = the number of studies.
*$p < .05$, ** $p < .01$, *** $p < .001$

Figure 1. Treatment Effectiveness by Year and Adherence to RNR Principles.



Discussion

Consistent with previous meta-analyses, the sexual and general recidivism rates for the treated sexual offenders were lower than the rates observed for the comparison groups (based on unweighted averages, 10.9% versus 19.2% for sexual recidivism; 31.8% versus 48.3% for any recidivism). These numbers are very similar to those reported by Lösel and Schmucker (2005; sexual recidivism rates of 11.1% versus 17.5%, based on a weighted average from 74 comparisons). Hanson et al. reported similar numbers (2002; sexual recidivism rates of 12.3% and 16.8% - based on an unweighted average from 38 studies and general recidivism rates of 27.9% versus 39.2% - based on 30 studies). Although the violent recidivism rates were not significantly different for the treated offenders compared to the untreated offenders in this review, the effect was in the same direction (unweighted averages of 22.9% versus 32.0%, respectively).

Confidence in the findings, however, must be tempered by the observation that most studies used weak research designs. Although this meta-analysis excluded 105 studies that did not meet minimum levels of study quality, of the remaining 23 studies, 18 were rated as weak and five were rated as good according to the CODC Guidelines. The effects tended to be stronger in the weak research designs compared to the good research designs. Reviewers restricting themselves to the better quality, published studies (Borduin et al., 2009; Hanson, Broom, & Stephenson, 2004; Marques et al., 2005; McGrath, Hoke, & Vojtisek, 1998; Meyer & Romero, 1980) could reasonably conclude that there is no evidence that treatment is effective in reducing sexual offence recidivism.

The treatments examined in the better studies, however, were diverse. If there is anything to be learned from the broad debate over the effectiveness of correctional rehabilitation, it is that not all interventions reduce recidivism. Multiple reviews and meta-analyses with general offender samples have demonstrated that the interventions that are most likely to reduce recidivism are those that meaningfully engage higher risk offenders in the process of changing their criminogenic needs (or criminogenic factors) (Andrews & Bonta, 2006; Andrews & Dowden, 2006; Landenberger & Lipsey, 2005; D. B. Wilson et al., 2005). The current review found that the same principles are also relevant to the treatment of sexual offenders. The pattern of results was completely consistent with the direction predicted by the principles of Risk, Need, and Responsivity (Andrews et al., 1990; Bonta & Andrews, 2007) in both the full set of studies as well as in the better, published studies (these latter analyses were not reported).

The analyses based on the Risk principle, however, were not statistically significant in any of the analyses in the current review. The Risk principle is also the principle with the least influence on treatment effectiveness for general offenders. Andrews and Dowden (2006) found that the average effect of correctional treatment was only slightly larger for studies that examined higher risk offenders (phi = .10 in 256 studies) than for studies involving lower risk offenders (phi = .05 in 74 studies). Although this difference is similar to the (non-significant) differences observed in the current set of studies, the magnitude of these differences is sufficiently small as to be of little practical value in most settings.

Landenberger and Lipsey's (2005) meta-analysis found strong support for the Risk principle, but the method they used for classifying risk levels artificially inflated the association between risk and treatment outcome. Specifically, offenders were classified as high risk or low risk based on the observed recidivism rates of the comparison groups. To understand the problem with this approach, imagine two matched sets of independent, random numbers drawn from the same population. When the values in Set 1 are sorted from highest to lowest, the expected value of the second set of numbers does not change. Consequently,

the highest values in sorted column (Set 1) would be expected to be larger than the population mean whereas the expected value of corresponding numbers in unsorted column (Set 2) would remain the population mean.

In the current review (and previous reviews), the Risk principle has been coded based on the risk level of the offenders participating in specific treatment programs. This definition does not fully express the meaning of the Risk principle, which states that interventions should be proportional to the offenders' risk of recidivism (Andrews & Bonta, 2006; Bonta & Andrews, 2007). Given that high risk offenders would be expected to require more treatment than moderate or low risk offenders (Bourgon & Armstrong, 2005), the relationship between risk and treatment effectiveness is unlikely to be linear.

Consequently, it may be preferable to consider the Risk principle at a broader level of program design and implementation. Since the 1990s, for example, the Correctional Service of Canada (CSC) has streamed sexual offenders into low, moderate, or high intensity programs based on initial assessments of risk and needs (National Committee on Sex Offender Strategy, 1996). Low and moderate intensity programs would screen out high risk offenders, directing these offenders toward more intensive (and more appropriate) interventions elsewhere. Consequently, the system would be adhering to the Risk principle, even if the Risk principle would not apply to individual programs considered in isolation. Evaluating such broad applications of the Risk principle would require comparing a complete cohort of CSC offenders with a complete cohort of offenders from a different jurisdiction in which the intensity of treatment was not matched to risk.

A hopeful finding of the current review was that recent treatments showed stronger treatment effects than older treatments. This finding is consistent with Hanson et al. (2002), but different from Lösel and Schmucker (2005), who found that the most effective programs were delivered in the 1970s. The different findings can partially be attributed to different selection criteria, and partially to the increasingly positive findings in the studies available since 2003 (the end date for Lösel and Schmucker's 2005 review).

For non-sexual offenders, treatment in the community has generally been more effective than treatment in institutions (Andrews & Bonta, 2006). The same pattern has not been found for sexual offenders in the current review, or in previous reviews (Hanson et al., 2002; Lösel & Schmucker, 2005). The sexual offender literature does not provide strong tests of whether program location matters given that no studies have directly compared the same treatment in both settings.

Previous reviews have found that treatment appeared to be equally effective for adult and adolescent sex offenders (Hanson et al., 2002; Lösel & Schmucker, 2005). The current review also found that the treatments offered to adult sexual offenders and the treatment offered to adolescent sexual offenders had similar overall effects. There was some difference in the general recidivism rates, but this appeared more related to the treatments given than to the age of the sample. In particular, two studies have found unusually strong effects of Multi-Systemic Therapy (MST) on the general recidivism rate of adolescent sex offenders (Borduin et al., 1990; Borduin et al., 2009).

Although it is not obvious how to generalize MST for adult sexual offenders, it was important to include MST studies in the current test of the RNR principles. Of the treatments examined, MST provided a rare example of a treatment consistent with all three principles of effective corrections, and was the only example of RNR-consistent treatment that was evaluated using a good research design.

Implications for Treatment Providers

We believe that the research evidence supporting the RNR principles is sufficient so that they should be a primary consideration in the design and implementation of intervention programs for sexual offenders. Evidence for the RNR principles is drawn from the current review and from the larger literature on effective correctional treatments (Andrews & Bonta, 2006). Most contemporary programs for sexual offenders already conform to some aspects of the Responsivity principle. Cognitive-behavioural treatments are the norm (McGrath, Cumming, & Burchard, 2003), and in the current review many of the programs examined also made special efforts to engage sexual offenders in treatment. Further research is needed concerning how best to apply the Risk principle to sexual offenders. Minimally, treatment providers should be cognisant that noticeable reductions in recidivism are not to be expected among the lowest risk offenders. Other treatment goals, such as meaningful re-integration into the community, may be appropriate for these cases.

Of the three RNR principles, attention to the Need principle would motivate the largest changes in the interventions currently given to sexual offenders. Much remains to be known about the criminogenic needs of sexual offenders; nevertheless, an empirical association with recidivism is a minimum criterion for a factor to be considered a potential dynamic risk factor (criminogenic need) (Kraemer et al., 1997). Many of the factors targeted in contemporary treatment programs do not meet this test. Offence responsibility, social skills training, and victim empathy are targets in more than 80% of sexual offender treatment programs (McGrath et al., 2003), yet none of these have been found to predict sexual recidivism (Hanson & Morton-Bourgon, 2004, 2005).

Consequently, it would be beneficial for treatment providers to carefully review their programs to ensure that the treatment targets emphasized are those empirically linked to sexual recidivism. Examples of promising criminogenic needs include sexual deviancy, sexual pre-occupation, low self-control, grievance thinking, and lack of meaningful intimate relationships with adults (Hanson & Morton-Bourgon, 2004, 2005).

Implications for Researchers

The use of the CODC Guidelines to assess study quality revealed that much can be done to reduce the bias and increase the confidence in outcome studies on sexual offender treatment. Many improvements to study quality can be done with limited cost, such as reporting intent-to-treat analyses, using equal and fixed follow-up periods, scoring actuarial risk measures on the treatment and comparison groups, using statistical controls, and matching on risk-relevant variables (see CODC, 2007b).

Although recidivism is the preferred outcome measure, it is also possible for researchers to examine short- and medium-term changes on intermediate targets and criminogenic needs. For example, Letourneau et al. (in press) used a strong research design to examine the effectiveness of treatment for adolescent sexual offenders in reducing: problematic sexual behaviour, delinquency, substance abuse, externalizing symptoms, and out-of-home placements. Improvements in such factors have inherent value, independent of their association with subsequent recidivism. Furthermore, they can be assessed using much shorter time periods (e.g., one year) than that required to assess sexual recidivism (three to five years minimum).

Further research is needed to identify the most important treatment targets for sexual offenders. In the field of general correctional treatment, the major criminogenic needs are well established (see meta-analytic reviews by Andrews & Bonta, 2006, chap. 13; Gendreau et al., 1996). Importantly, programs that deliberately target the central criminogenic needs, such as criminal attitudes (Andrews, 1980) and substance abuse (Gottfredson, Najaka, Kearley, & Rocha, 2006) have been shown to reduce general recidivism rates.

In contrast, much less is known about the processes by which sexual offenders change. The current review found a general, overall pattern that indicated programs emphasizing risk factors empirically associated with recidivism were more effective than programs that emphasized other factors. It is not uncommon, however, for research to find that improvements on factors presumed to be criminogenic do not influence sexual recidivism rates (Langton, 2003; Quinsey, Khanna, & Malcolm, 1998; however, see Olver, Wong, & Nicholaichuck, 2007, for an exception). Outcome research can help advance knowledge of the change process by routinely reporting the relationship between changes on intermediate targets (i.e., criminogenic needs) and subsequent recidivism.

Finally, strong studies are needed. Of the 129 studies of treatment for adult or adolescent sexual offenders examined using the CODC Guidelines, 19 were rated as weak, 5 were good, and 81% (105) were rejected. None were rated as strong. In agreement with Seto et al. (2008), we believe that an important requirement of strong research designs is the experimenter's ability to determine subject assignment based on a procedure that controls for both measured and unmeasured features of the offenders (i.e., random assignment). Although random assignment studies are difficult to implement and are unpopular with certain groups (e.g., Marshall & Marshall, 2007), they can be done with sexual offenders (Borduin et al., 1990, 2009; Letourneau et al., in press; Marques et al., 2005; Meyer & Romero, 1980; Robinson, 1995). These studies remain the best available alternative for minimizing subject selection bias. Random selection is also one of the most ethically defensible methods of assigning individuals to treatment when demand exceeds supply, or the relative superiority of alternate treatments is unknown. Readers sympathetic to sexual offender rehabilitation may be content with the encouraging findings from weak research designs; however, sceptics will only be compelled to change their opinions by the strongest possible evidence.

References

References marked with an asterisk were included in the meta-analysis.

Abracen, J., Looman, J., & Nicholaichuk, T. P. (1999). *Recidivism among treated sexual offenders and matched comparison subjects: Data from the Regional Treatment Centre (Ontario) post-1989 sample.* Unpublished manuscript.

Andrews, D. A. (1980). Some experimental investigations of the principles of differential association through deliberate manipulations of the structure of service systems. *American Sociological Review, 45*, 448-462.

Andrews, D. A., & Bonta, J. (2006). *The psychology of criminal conduct* (4th ed.). Newark, NJ: LexisNexis/Anderson.

Andrews, D. A., Bonta, J., & Hoge, R. D. (1990). Classification for effective rehabilitation: Rediscovering psychology. *Criminal Justice and Behavior, 17*, 19-52.

Andrews, D. A., & Dowden, C. (2006). Risk principle of case classification in correctional treatment. *International Journal of Offender Therapy and Comparative Criminology, 50,* 88-100.

Aos, S., Phipps, P., Barnoski, R., & Lieb, R. (1999). *The comparative costs and benefits of programs to reduce crime: A review of national research findings with implications for Washington State* (Document No. 99-05-1202). Olympia, WA: Washington State Institute for Public Policy.

*Bakker, L., Hudson, S., Wales, D., & Riley, D. (1998). *And there was light: Evaluating the Kia Marama treatment programme for New Zealand sex offenders against children.* Christchurch: New Zealand Department of Corrections.

Bonta, J., & Andrews, D. A. (2007). *Risk-need-responsivity model for offender assessment and rehabilitation* (Corrections Research User Report No. 2007-06). Ottawa, Ontario: Public Safety Canada.

*Borduin, C. M., Henggeler, S. W., Blaske, D. M., & Stein, R. J. (1990). Multisystemic treatment of adolescent sexual offenders. *International Journal of Offender Therapy and Comparative Criminology, 34,* 105-113.

*Borduin, C. M., Schaeffer, C. M., & Heiblum, N. (2009). A randomized clinical trial of multisystemic therapy with juvenile sexual offenders: Effects on youth social ecology and criminal activity. *Journal of Consulting and Clinical Psychology, 77,* 26-37.

Borenstein M., Hedges L., Higgins J., & Rothstein, H. (2005). *Comprehensive Meta-analysis Biostat* (Version 2.0) [Computer software]. Englewood, NJ.

Bourgon, G., & Armstrong, B. (2005). Transferring the principles of effective treatment into a 'real world' prison setting. *Criminal Justice and Behavior, 32,* 3-25.

Campbell, M. A., French, S., & Gendreau, P. (2007). *Assessing the utility of risk assessment tools and personality measures in the prediction of violent recidivism for adult offenders* (Corrections Research User Report No. 2007-04). Ottawa, Ontario: Public Safety Canada.

Carpentier, M. Y., Silovsky, J. F., & Chaffin, M. (2006). Randomized trial of treatment for children with sexual behavior problems: Ten-year follow-up. *Journal of Consulting and Clinical Psychology, 74*, 482-488.

Collaborative Outcome Data Committee. (2007a). *Sexual offender treatment outcome research: CODC Guidelines for evaluation Part 1: Introduction and overview* (Corrections Research User Report No. 2007-02). Ottawa, Ontario: Public Safety Canada.

Collaborative Outcome Data Committee. (2007b). *The Collaborative Outcome Data Committee's guidelines for the evaluation of sexual offender treatment outcome research Part 2: CODC guidelines* (Corrections Research User Report No. 2007-03). Ottawa, Ontario: Public Safety Canada.

*Cooper, H. M. (2000). Long-term follow-up of a community-based treatment program for adolescent sex offenders. *Masters Abstracts International, 45* (03). (UMI No. MR21542).

Craissati, J., & Beech, A. R. (2005). Risk prediction and failure in a complete urban sample of sex offenders. *Journal of Forensic Psychiatry and Psychology,* 16, 24-40.

*Davidson, P. R. (1984, March). *Behavioral treatment for incarcerated sex offenders: Post-release outcome.* Paper presented at A Conference on Sex Offender Assessment and Treatment, Kingston, Ontario, Canada.

Fleiss, J. L. (1994). Measures of effect size for categorical data. In H. Cooper and L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 245-260). New York: Russell Sage.

*Friendship, C., Mann, R. E., & Beech, A. R. (2003). Evaluation of a national prison-based treatment program for sexual offenders in England and Wales. *Journal of Interpersonal Violence, 18*, 744-759.

Furby, L., Weinrott, M. R., Blackshaw, L. (1989). Sex offender recidivism: A review. *Psychological Bulletin, 105*, 3-30.

Gallagher, C. A., Wilson, D. B., Hirschfield, P., Coggeshall, M. B., & MacKenzie, D. L. (1999). A quantitative review of the effects of sex offender treatment on sexual reoffending. *Corrections Management Quarterly, 3*, 19-29.

Gendreau, P., Little, T., & Goggin, C. (1996). A meta-analysis of the predictors of adult offender recidivism: What works! *Criminology, 34*, 575-607.

Gottfredson, D. E., Najaka, S. S., Kearley, B. W., & Rocha, C. M. (2006). Long-term effects of participation in the Baltimore City drug treatment court: Results from an experimental study. *Journal of Experimental Criminology, 2,* 67-98.

Gutierrez, L. (2008). *A meta-analysis of drug treatment court literature: Assessing study and treatment quality*. Unpublished undergraduate's thesis, Carleton University, Ottawa, Ontario, Canada.

Hall, G. C. N. (1995). Sexual offender recidivism revisited: A meta-analysis of recent treatment studies. *Journal of Consulting and Clinical Psychology, 63*, 802-809.

*Hanson, R. K., Broom, I., & Stephenson, M. (2004). Evaluating community sex offender treatment programs: A 12-year follow-up of 724 offenders. *Canadian Journal of Behavioural Science, 36*, 87-96.

Hanson, R. K., Gordon, A., Harris, A. J. R., Marques, J. K., Murphy, W., Quinsey, V. L., & Seto, M. C. (2002). First report of the Collaborative Outcome Data Project on the effectiveness of psychological treatment of sex offenders. *Sexual Abuse: A Journal of Research and Treatment, 14*, 169-194.

Hanson, R. K. & Morton-Bourgon, K. (2004). *Predictors of sexual recidivism: An updated meta-analysis* (Corrections Research User Report No. 2004-02). Ottawa, Ontario: Public Safety Canada.

Hanson, R. K., & Morton-Bourgon, K. E. (2005). The characteristics of persistent sexual offenders: A meta-analysis of recidivism studies. *Journal of Consulting and Clinical Psychology, 73*, 1154-1163.

*Hanson, R. K., Steffy, R. A., & Gauthier, R. (1993). Long-term recidivism of child molesters. *Journal of Consulting and Clinical Psychology, 61*, 646-652.

*Harkins, L. (2004). Recidivism and within-treatment change among treated sex offenders and matched comparison subjects. *Masters Abstracts International, 43* (03), p. 993. (UMI No. MQ95168).

Harris, A., Phenix, A., Hanson, R. K., & Thornton, D. (2003). *Static-99 coding rules: Revised 2003*. Ottawa, ON: Solicitor General Canada.

Harris, G. T., Rice, M. E., & Quinsey, V. L. (1998). Appraisal and management of risk in sexual aggression: Implications for criminal justice policy. *Psychology, Public Policy, and Law, 4*, 73-115.

Hedges, L. V. (1994). Fixed effect models. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 285-299). New York: Russell Sage.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. New York: Academic Press.

Hedges, L. V., & Vevea, J. L. (1998). Fixed- and random-effects models in meta-analysis. *Psychological Methods, 3*, 486-504.

Helmus, L. (2008). Structured guidelines for evaluating study quality. In G. Bourgon, R. K. Hanson, J. D. Pozzulo, K. E. Morton-Bourgon, & C. L. Tanasichuk, (Eds.) *The Proceedings of the 2007 North American Correctional & Criminal Justice Psychology Conference* (pp. 45-49). (Corrections User Report 2008-02). Ottawa, Ontario: Public Safety Canada.

Juni, P., Witschi, A., Bloch, R., & Egger, M. (1999). The hazards of scoring the quality of clinical trials for meta-analysis. *Journal of the American Medical Association, 282*, 1054-1060.

Kenworthy, T., Adams, C. E., Brooks-Gordon, B., & Fenton, M. (2004). *Psychological interventions for those who have sexually offended or are at risk of offending* (CD004858). Cochrane Database of Systematic Reviews, Issue 3. Chichester, UK: John Wiley & Sons.

Kraemer, H. C., Kazdin, A. E., Offord, D. R., Kessler, R. C., Jensen, P. S., & Kupler, D. J. (1997). Coming to terms with the terms of risk. *Archives of General Psychiatry, 54,* 337-343.

Kunz, R., & Oxman, A. (1998). The unpredictability paradox: Review of empirical comparisons of randomized and non-randomized clinical trials. *British Medical Journal, 317,* 1185-90.

Landenberger, N. A., & Lipsey, M. W. (2005). The positive effects of cognitive-behavioral programs for offenders: A meta-analysis of factors associated with effective treatment. *Journal of Experimental Criminology, 1,* 451-476.

Langton, C. M. (2003). Contrasting approaches to risk assessment with adult male sexual offenders: An evaluation of recidivism prediction schemes and the utility of supplementary clinical information for enhancing predictive accuracy. *Dissertations Abstracts International, 64,* 1907B. (UMI No. NQ78052).

Letourneau, E. J., Henggeler, S. W., Borduin, C. M., Schewe, P. A., McCart, M. R., Chapman, J. E., & Saldana, L. (in press). Multisystemic Therapy for juvenile sexual offenders: 1-year results from a randomized effectiveness trial. *Journal of Family Psychology.*

Lösel, F., & Schmucker, M. (2005). The effectiveness of treatment for sexual offenders: A comprehensive meta-analysis. *Journal of Experimental Criminology, 1,* 117-146.

*Malcolm, P. B., Marshall, L., Marshall, W. L. (2004). *Outcome of a motivational preparatory program for sexual offenders: A comparison with a matched control.* Unpublished manuscript. [Rockwood].

Mander, A. M., Atrops, M. E., Barnes, A. R., & Munafo, R. (1996). *Sex offender treatment program: Initial recidivism study.* Anchorage, Alaska: Offender Program, Alaska Department of Corrections and Alaska Justice Statistical Analysis Unit, Justice Center, University of Alaska, Anchorage.

*Marques, J. K., Wiederanders, M., Day, D. M., Nelson, C., & van Ommeren, A. (2005). Effects of a relapse prevention program on sexual recidivism: Final results from California's Sex Offender Treatment and Evaluation Project (SOTEP). *Sexual Abuse: A Journal of Research and Treatment, 17,* 79-107.

*Marshall, W. L., Eccles, A., & Barbaree, H. E. (1991). The treatment of exhibitionists: A focus on sexual deviance versus cognitive and relationship features. *Behaviour Research and Therapy, 29,* 129-135.

Marshall, L. E., & Marshall, W. L. (2007). The utility of the random controlled trial for evaluating sexual offender treatment: The gold standard or an inappropriate strategy? *Sexual Abuse: A Journal of Research and Treatment, 19,* 175-191.

*Marshall, L. E., Marshall, W. L., Fernandez, Y. M., Malcolm, P. B., & Moulden, H. M. (2008). The Rockwood Preparatory Program for sexual offenders: Description and preliminary appraisal. *Sexual Abuse: A Journal of Research and Treatment, 20,* 25-42. [Rockwood].

*Martin, I. (1998). Efficacité d'un programme cognitif-behavioural institutionnel pour délinquants sexuels [Effectiveness of an institutional cognitive-behavioural program for sexual offenders]. *Dissertation Abstracts International, 61* (01), 518B. (UMI No. NQ43732).

Martinson, R. (1974). What works? – Questions and answers about prison reform. *The Public Interest, 35,* 22-54.

McGrath, R. J., Cumming, G. F., & Burchard, B. L. (2003). *Current practices and trends in sexual abuser management: The Safer Society 2002 nationwide survey*. Brandon, VT: Safer Society Press.

*McGrath, R. J., Hoke, S. E., & Vojtisek, J. E. (1998). Cognitive-behavioral treatment of sex offenders. *Criminal Justice and Behavior, 25*, 203-225.

*Meyer, L. C., & Romero, J. (1980). *A ten-year follow-up of sex offender recidivism*. Philadelphia, PA: Pennsylvania Commission on Crime and Delinquency.

Moher, D., Jadad, A. R., Nichol, G., Penman, M., Tugwell, P., & Walsh, S. (1995). Assessing the quality of randomized controlled trials: An annotated bibliography of scales and checklists. *Controlled Clinical Trials, 16*, 62-73.

*Nathan, L., Wilson, N. J., & Hillman, D. (2003). *Te Whakakotahitanga: An evaluation of the Te Piriti Special Treatment Programme for child sex offenders in New Zealand* (ISBN 047825201). New Zealand: Department of Corrections.

National Committee on Sex Offender Strategy. (1996). *Standards and guidelines for the provision of services to sex offenders*. Ottawa, Ontario: Correctional Service of Canada. Retrieved November 26, 2008, from http://198.103.98.138/text/pblct/so/standards/stande-eng.shtml.

Olver, M. E., Wong, S. C. P., Nicholaichuk, T., & Gordon, A. (2007). The validity and reliability of the Violence Risk Scale – Sexual Offender Version: Assessing sex offender risk and evaluating therapeutic change. *Psychological Assessment, 19*, 318-329.

Overton, R. C. (1998). A comparison of fixed-effects and mixed (random-effects) models for meta-analysis tests of moderator variable effects. *Psychological Methods, 3,* 354-379.

*Proctor, E. (1996). A five-year outcome evaluation of a community-based treatment program for convicted sexual offenders run by the probation service. *Journal of Sexual Aggression, 2,* 3-16.

Quinsey, V. L., Khanna, A., & Malcolm, P. B. (1998). A retrospective evaluation of the Regional Treatment Centre sex offender treatment program. *Journal of Interpersonal Violence, 13*, 621-644.

Rice, M. E., & Harris, G. T. (2003). The size and sign of treatment effects in sex offender therapy. *Annals of the New York Academy of Sciences, 989*, 428-440.

*Robinson, D. (1995). *The impact of cognitive skills training on post-release recidivism among Canadian federal offenders* (User Report R-41). Ottawa, Ontario: Correctional Service Canada.

*Ruddijs, F., & Timmerman, H. (2000). The Stichting Ambulante Preventie ProjectenMethod: A comparative study of recidivism in first offenders in a Dutch outpatient setting. *International Journal of Offender Therapy and Comparative Criminology, 44*, 725-739.

Seto, M. C., Marques, J. K., Harris, G. T., Chaffin, M., Lalumière, M. L., Miner, M. H., Berliner, L., Rice, M. E., Lieb, R., & Quinsey, V. (2008). Good science and progress in sex offender treatment are intertwined: A response to Marshall and Marshall (2007). *Sexual Abuse: A Journal of Research and Treatment, 20,* 247-255.

Sherman, L. W., Gottfredson, D., Mackenzie, D., Eck, J., Reuter, P., & Bushway, S. (1997). *Preventing crime: What works, what doesn't, what's promising. A report to the United States Congress.* College Park, Maryland: University of Maryland, Department of Criminology and Criminal Justice.

Simpson, B., Bonta, J., Bourgon, G., Rugge, T., Yessine, A., Helmus, L., & Gutierrez, L. (2008, June). What we know and where we need to go for effective community supervision. In G. Bourgon, L. Gutierrez, T. Rugge, & K. Simpson (Chairs), *From "what works" to "making it work": The strategic training initiative in community supervision (STICS) project.* Symposium conducted at the 69th Annual Convention of the Canadian Psychological Association, Halifax, Nova Scotia, Canada.

*Taylor, R. (2000). A seven-year reconviction study of HMP Grendon therapeutic community. *Research, Development and Statistics Directorate Research Findings, 115*, 1-4. London: Home Office.

*Ternowski, D. R. (2004). *Sex offender treatment: An evaluation of the Stave Lake Correctional Centre Program*. Dissertation Abstracts International, 66 (06), 3428B. (UMI No. NR03201).

Whitehead, A. (2002). *Meta-analysis of controlled clinical trials*. Chichester, UK: John Wiley & Sons.

Wilson, D. B., Bouffard, L.A., & Mackenzie, D. L. (2005). A quantitative review of structured, group-oriented, cognitive-behavioral programs for offenders. *Criminal Justice and Behavior, 32*, 172-204.

*Wilson, R. J., Picheca, J. E., & Prinzo, M. (2005). *Circles of Support & Accountability: An evaluation of the pilot project in south-central Ontario* (User Report R-168). Ottawa, Ontario: Correctional Service Canada.

*Worling, J. R., & Curwen, T. (2000). Adolescent sexual offender recidivism: Success of specialized treatment and implications for risk prediction. *Child Abuse & Neglect, 24*, 965-982.